DOCUMENT RESUME

ED 049 316	241	TM 000 517

AUTHOR	Borich, Gary D.
TITLE	Homogeneity of Slopes Test for Multiple Regression Equations with Reference to Aptitude-Treatment Interactions.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	Feb 71
GRANT	OEG-5-70-0008-010; OEG-6-0-9-247053-3589
NOTE	14p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Analysis of Covariance, Correlation, *Interaction, *Interaction Process Analysis, *Mathematical Models, *Multiple Regression Analysis, Predictor Variables, Tests of Significance

ABSTRACT

Methods exist for testing the homogeneity of group regression for the case in which there is only one predictor. Studies which have investigated aptitude-treatment interactions have adopted the homogeneity of regressions test as standard methodology for assessing the difference in regression slopes across treatments. Additional statistical methodology is suggested by which the homogeneity of group regressions can be tested when two or more predictors are present. Before covariate analysis can be used, the homogeneity of group regression must be tested. To test the hypothesis that the slopes are equal for the homogeneity of group regressions for a single covariate, one can follow the standard linear prediction model. For the case in which there are multiple covariates, a test for homogeneity of regressions provides an overall estimate of the difference between treatments, taking into account the effect of the multiple covariates upon each treatment, simultaneously. Partial hypotheses for a multiple covariate model are presented in order to isolate the cause of overall interaction when treatment slopes with multiple covariates are not equal. Data from two aptitude-treatment investigations are used to illustrate the equations. (CK)

ERIC Full Text Provided by ERIC

Institute For Child Study

School of Education   Indiana University

ERIC Full Text Provided by ERIC

Homogeneity of Slopes Test for Multiple Regression Equations with Reference to Aptitude-Treatment Interactions

Gary D. Borich

Institute for Child Study, Indiana University

Paper presented at the annual meeting of the American Educational Research Association, February 4th, 1971, in New York City.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

The author would like to thank Salis Halperin, Syracuse University and Richard Snow, Stanford Center for Research and Development in Teaching for reading an earlier draft of this paper and for their helpful suggestions.

The research reported herein was supported by U.S.O.E. grants No. G-5-70-0008-010 and OE6-0-9-247053-3589.

Homogeneity of Slopes Test for Multiple Regression

Equations with Reference to Aptitude-Treatment

Interactions

Gary D. Borich

Institute for Child Study, Indiana University

Walker and Lev (1953) and Edwards (1968) illustrate a method for testing the homogeneity of group regressions for the case in which there is one predictor. Studies which have investigated aptitude-treatment interactions (see Cronbach and Snow, 1969) have adopted the homogeneity of regressions test as standard methodology for assessing the difference in regression slopes across treatments. The statistical model for this test, however, is inappropriate for the case in which there are two or more aptitude variables. The purpose of this paper is to suggest additional statistical methodology by which the homogeneity of group regressions can be tested when two or more predictors are present.

The analysis of covariance model. The homogeneity of group regressions test may be familiar to the reader as the test which precedes analysis of covariance. The purpose of the test is to determine whether or not regressions of the dependent measure on the covariate differ significantly across treatments. An underlying assumption of analysis of covariance is not met when regressions significantly differ.

The homogeneity of regression lines test (Walker and Lev, 1953; Edwards, 1968) is performed with one covariate and one criterion for

ERIC Full Text Provided by ERIC

multiple treatment groups. The F ratio for the homogeneity of regression lines test is derived from the variation of two sources: (a) observations within each treatment group about the regression for the groups, and (b) observations within each treatment group about the regression lines with a common slope. The error term is represented by (a), while the difference between (a) and (b) represents the treatment variation. A brief review of the homogeneity of regression lines test will be used to illustrate the general model. After which, we shall extend the model to test for homogeneity of regressions when multiple covariates are present.*

Homogeneity of group regressions, single covariate. To test the hypothesis that $B_1 = B_2 = \ldots = B_k = B$ (i.e. the slopes are equal), we start with the standard linear prediction model:

$$Y_{ij} = a_j + B_j X_{ij} + e_{ij};\ j = 1, \ldots, k;\ i = 1, \ldots, n_j$$

where Y_{ij} is the criterion, a_j is the intercept of the j^{th} group, B_j is the slope in the j^{th} group, X_{ij} is the covariate, k is the number of groups, and n_j is the number of subjects in the j^{th} group.

The residual sum of squares (i.e., $\Sigma \hat{e}_{ij}^2$) has degrees of freedom given by the number of subjects minus the number of parameters fit. Therefore, we have N-$2k$ degrees of freedom.

---

*While analysis of covariance can be performed with various computer programs, these programs do not commonly test for homogeneity of regressions. Multiple regression programs, including BMD03R, however, may be used to obtain the quantities specified in this paper.

To test that $B_1 = B_2 = \ldots = B_k = B$ we next fit the data to a second more restrictive model (observations within each treatment group about the regression lines with a common slope) given by:

$$Y_{ij} = a_j + BX_j + f_{ij};\ j = 1, \ldots, k;\ i = 1, \ldots, n_j$$

For the residual sum of squares ($\Sigma\hat{f}^2_{ij}$) we have N-k-1 degrees of freedom.

Since the restricted model combines treatment groups, we expect $\Sigma\hat{f}^2_{ij}$ to be greater than $\Sigma\hat{e}^2_{ij}$. These can be equal if the hypothesis is true, but $\Sigma\hat{f}^2_{ij}$ can not be less than $\Sigma\hat{e}^2_{ij}$.

To test for equal slopes, we form a hypothesis sum of squares given by $SS_{hyp} = \Sigma\hat{f}^2_{ij} - \Sigma\hat{e}^2_{ij}$ with (N-k-1) - (N-2k) = k-1 degrees of freedom. An $\underline{F}$ test can then be formed utilizing observations within each group about the regression for the group as an estimate of error:

$$F\ (k-1,\ N-2k) = \frac{SS_{hyp}\ /\ (k-1)}{\Sigma\hat{e}^2_{ij}\ /\ (N-2k)}$$

To the extent that covariates are unrelated (r = .00) in multiple covariate problems there is justification for performing the homogeneity of regression lines test separately for each covariate. When such a relationship is not obtained, we must take into account the relation between covariates. Other models in which a null relationship between covariates is not assumed are more generally applicable to multiple covariate problems.

Homogeneity of group regressions, multiple covariates. For the case in which there are multiple covariates, a test between hyper planes is analogous to the

Walker and Lev and Edwards test of regression lines. A test for homogeneity of regressions with multiple covariates provides an overall estimate of the difference between treatments, taking into account the effect of the multiple covariates upon each treatment simultaneously. The error term for such a test is given by the summed residual sum of squares for treatments, while the treatment variation is given by the summed residual sum of squares for treatments minus the residual sum of squares for the treatment groups combined.

If we have n number of covariates, $X_1, X_2, \ldots, X_n$, our full model becomes:

$$Y_{ij} = a_j + B_{1j}X_{1ij} + B_{2j}X_{2ij} +, \ldots, B_{nj}X_{nij} + e_{ij}$$

with N-pk degrees of freedom, where p equals the number of parameters fit.

Constructing the restricted model for multiple covariates we have:

$$Y_{ij} = a_j + B_1X_{1ij} + B_2X_{2ij} + , \ldots, B_n X_{nij}f_{ij}$$

with N-k-(p-1) degrees of freedom, where p-1 equals the number of covariates.

The sum of squares for the hypothesis of equal slopes is given by

$$SS_{hyp} = \Sigma \hat{f}^2_{ij} - \Sigma \hat{e}^2_{ij}$$

which has (N-k-(p-1) - (N-pk) = p(k-1) degrees of freedom.

The F test, again utilizing observations within each group about the regression for the group as an estimate of error, is given by

$$F_{p(k-1), (N-pk)} = \frac{SS_{hyp} \;/\; p(k-1)}{\Sigma \hat{e}^2_{ij} \;/\; (N-pk)}$$

The homogeneity of regressions test with multiple covariates measures the overall treatment effect but does not indicate differences in the regressions which may be due to any one covariate. The result, therefore, is a generalized test which determines the significance of the treatment variation but not the separate effect of the covariates upon the treatment variation.

Partial hypotheses for the multiple covariate model. In order to isolate the cause of the overall interaction we can construct partial hypotheses based upon the restricted model. Here, we make no assumptions of uncorrelated covariates as would be the case if the regression lines test were applied. To form partial hypotheses we construct the model:

$$Y_{ij} = a_j + B_1 X_{1ij} + B_{2j} X_{2ij} + B_{nj} X_{nij} + e_{ij}$$

in which treatments are combined for one covariate and allowed to differ for the remaining covariates. For the partial hypothesis sum of squares we have $\Sigma \hat{g}^2_{ij} - \Sigma \hat{e}^2_{ij}$ with $(N-(p-1)k-1-(N-pk) = (k-1)$ degrees of freedom. The F test for this hypothesis is given by:

$$F(k-1, N-3k) = \frac{SS_{hyp} / (k-1)}{\Sigma \hat{e}^2_{ij} / (N-pk)}$$

Partial hypothesis are constructed for each covariate to identify the causes of the interaction. For each covariate tested, a common slope is formed, while slopes for all other covariates are allowed to differ by treatment.

Example data: high and low correlation between aptitudes. Data from two aptitude-treatment investigations are used to illustrate the

foregoing equations. While both investigations manipulated two treatments and obtained measures for two aptitudes, one reports a low intercorrelation between aptitudes (Koran 1969) and the other reports a high intercorrelation between aptitudes (Borich 1970). For the former study the correlation between aptitudes for treatment 1 was -.11 and for treatment 2, -.12, while for the latter study correlations between aptitudes for the two treatments were -.82 and -.55. Response surfaces for the studies appear in Figures 1 and 2.

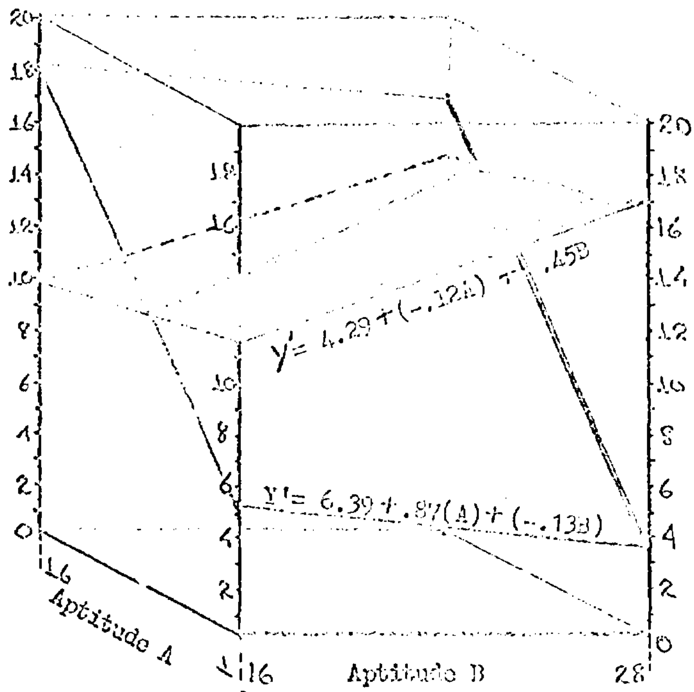

FIG. 1. [illegible] Data

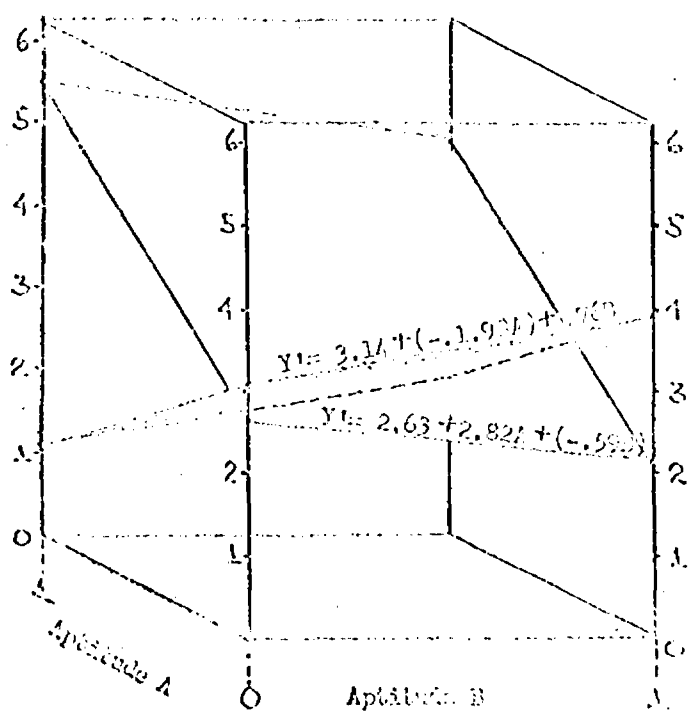

FIG. 2. [illegible]

ERIC Full Text Provided by ERIC

We may apply the homogeneity of regressions test for multiple covariates to each investigation. For the Koran data represented in Fig. 1, we have,

$$\text{Homogeneity of Multiple Regressions} = \frac{3451.29 - (1345.47 + 1538.91)/2}{1345.47 + 1538.91 / (36 + 40 - 6)} = 6.88 \text{ (df 2, 70)}$$

And, for the Borich data represented in Fig. 2, we have,

$$\text{Homogeneity of Multiple Regressions} = \frac{43.96 - (16.76 + 12.12)/2}{16.76 + 12.12 / (15 + 15 - 6)} = 6.13 \text{ (df 2, 24)}$$

Treatment differences for each covariate may be determined with partial hypotheses (i.e. $B_{1j} = B_1$). To isolate the cause of the interactions we construct all possible partial hypotheses. For the above data these are:

$$B_1 = B_{1j} \quad (B_{2j}\text{'s Differ})$$

$$B_2 = B_{2j} \quad (B_{1j}\text{'s Differ})$$

Identical models are applied to each investigation in order to test the partial hypotheses. For the partial hypotheses we construct the restricted models.

$Y_{ij} = a_j + B_1X_{1ij} + B_{2j}X_{2ij} + e_{ij}$ (to test $B_1 = B_{1j}$)
with df given by $N - k - 1 - k = N - 2k - 1$, and

$Y_{ij} = a_j + B_{1j}X_{1ij} + B_2X_{2ij} + e_{ij}$ (to test $B_2 = B_{2j}$)
with df given by $N - k - k - 1 = N - 2k - 1$, with the full model given by

$Y_{ij} = a_j + B_{1j}X_{1ij} + B_{2j}X_{2ij} + e_{ij}$
with $N - k - k - k = N - 3k$ degrees of freedom.

In order to calculate F ratios for the partial hypotheses, treatments are dummy or contrast coded and placed in the regression equations above. Further examples with dummy and contrast codes are provided by Cohen (1968), Bottenberg and Ward (1963), and Hamilton (1969) and need not be repeated here.

For the above investigations it is instructive to note the effect of the strong and weak interrelationships between covariates upon the regression lines test.

By determining the sum of squares for each covariate and comparing their sum to the regression sum of squares for treatments, we may identify the extent to which the covariates, separately, can account for treatment variation when the covariates are in a multiple regression equation. When the summed sum of squares for each covariate equals the between treatment variation, the correlation between covariates is zero. Or,

$$\frac{\Sigma SS_{c_i}}{\Sigma SS_{r_j} - SS_r} = 1$$

where ΣSS_{c_i} is the covariate sum of squares for i covariates, SS_{r_j} the regression sum of squares for j groups, and SS_r the combined treatment regression sum of squares. The difference between the summed sum of squares for covariates and the regression sum of squares for treatments will increase as the correlation between covariates increases. Therefore, the percent of the treatment variation due to regression that can be accounted for by the sum of squares of the variables separately can be determined. For the Borich data,

Aptitude intercorrelations by treatments:

$$\frac{SS_{c_1} + SS_{c_2} = 5.78}{SS_{r_1} + SS_{r_2} - SS_r = 14.94}$$

(r=.-55, r-.82)

And for the Koran data,

$$\frac{SS_{v_1} + SS_{v_2} = 312.93}{SS_{r_1} + SS_{r_2} - SS_r = 385.80}$$

(r=-.11, r-.12)

ERIC

In the former study we can note that the sum of squares for the aptitudes separately, accounts for only 39 percent of the treatment variation, while for the Koran data the sum of squares for the aptitude, separately, accounts for 81 percent of the treatment variation. As the interrelationship between aptitudes increases, the difference between the aptitude sum of squares and the regression sum of squares for treatments increases. For the Borich data, the Walker and Lev model would fail to consider the strong interrelationship between aptitudes, while for the Koran data the omission is not as great. The percent of treatment variation that can be accounted for by the aptitudes separately should be reported when an interrelationship between aptitudes exists and when the homogeneity of regression lines test used. However, no assumption as to the interrelationship of aptitudes need be made when the partial hypothesis test ($B_{1j} = B_1$) is employed. The partial hypothesis model is applicable when overall treatment slopes with multiple covariates are not equal and when the specific covariates causing the interaction are to be identified.

References

Borich, G. D. Final Report: Learning and transfer in concept attainment as a function of concept rule, cue saliency, and stimulus variety. Bloomington, Ind.: Indiana University 1971.

Bottenberg, R.A. and Ward, J.H. Applied multiple linear regression (PRL-RDR-63-6), Lackland AF Base, Texas, 1963.

Cohen, J. Multiple regression as a general data-analytic system. Psychological Bulletin, 1968, 6, 426-443.

Cronback, L. J., & Snow, R.E. Final report: Individual differences in learning ability as a function of instructional variables. Stanford, California.: Stanford University, March 1969.

Edwards, A.L. Experimental design in psychological research (3rd ed.) New York: Holt, 1968.

Hamilton, N.R. Differential response to instruction designed to call upon spatial and verbal aptitudes. Technical report No. 5, Stanford Center for Research and Development in Teaching (1969).

Koran, M.R. The effects of individual differences on observational learning in the acquisition of a teaching skill. Unpublished doctoral dissertation, Stanford University, 1969.

Walker, H. and Lev, J. Statistical inference. New York: Holt, 1953.